Modélisation à partir d'Images

Edmond Boyer, Peter Sturm

Chapitre 1

Modélisation à partir d'images

Dans ce chapitre, nous nous intéressons à la modélisation d'objets et de scènes tridimensionnelles à partir d'images. La vision par ordinateur, domaine qui a beaucoup évolué ces dernières années, propose dorénavant de nombreuses approches permettant de reconstruire les éléments d'une scène à partir d'images (photographies) de cette dernière. Nous précisons ici le contexte théorique nécessaire pour comprendre ces différentes méthodes, en particulier les outils géométriques qui sont utilisés, puis nous introduisons les grands principes mis en œuvre par ces méthodes. Nous illustrons ces principes par la présentation de plusieurs méthodes populaires de modélisation à partir d'une ou de plusieurs images.

1.1. Introduction

La modélisation de scènes ou d'objets tridimensionnels à partir d'images appartient au domaine de la *vision par ordinateur*, domaine où l'on cherche à reproduire certaines fonctionnalités de la vision humaine. L'aspect qui nous intéresse dans ce document est la perception tridimensionnelle de l'environnement et, en particulier, la capacité de produire des modèles de cet environnement à partir d'images. Un des principaux domaines d'applications est la *réalité virtuelle*. L'acquisition de modèles constitue en effet une étape préliminaire nécessaire en réalité virtuelle. Les applications dans ce domaine bénéficient donc grandement de méthodes d'acquisitions automatiques ; ces dernières remplaçant une acquisition manuelle basée sur les modeleurs logiciels qui est souvent fastidieuse. De plus, la modélisation à partir d'images conduit à un réalisme avancé des modèles virtuels produits : le *photo-réalisme*, ceci grâce aux

Chapitre rédigé par Edmond Boyer et Peter Sturm.

techniques de *placage de texture*[1] qui utilisent l'information photométrique contenue dans les images. Les modèles photo-réalistes peuvent ensuite être utilisés pour créer des mondes virtuels, ou pour être ajoutés à d'autres images réelles dans des applications de *réalité augmentée*. Les applications de la modélisation à partir d'images sont nombreuses et variées. Pour exemple, citons les applications en géologie où un modèle de terrain est construit à partir d'images [aériennes], ainsi que les applications en architecture où un modèle photo-réaliste de bâtiments peut être obtenu à partir d'images, ce modèle pouvant ensuite être incrusté dans d'autres images d'environnements urbains. Citons aussi les applications construisant des modèles de visages, ou d'autres parties du corps humain, à partir d'images, et conduisant ainsi à la réalisation de clones humains virtuels.

Il s'agit donc de produire des modèles virtuels d'environnements, ou d'objets, à partir d'images de ces derniers. Une image est une projection d'une scène, généralement un espace à trois dimensions (largeur, hauteur et profondeur), sur un plan, soit un espace à deux dimensions (largeur et hauteur). Le problème qui se pose alors est de déterminer les caractéristiques tridimensionnelles qui constitueront le modèle à partir de projections planes, et donc bidimensionnelles, de la scène contenant le modèle. La vision par ordinateur propose plusieurs solutions pour résoudre ce problème. Ces solutions font bien entendu intervenir la géométrie et en particulier la *géométrie projective* qui offre un cadre mathématique facilitant la manipulation des projections. La géométrie constitue un élément fondamental de la modélisation à plusieurs titres. Il est en effet nécessaire de comprendre comment se forme une image et donc de quelle manière une scène est projetée sur une image, ou en d'autres termes sur un tableau de pixels. Il est aussi nécessaire de déterminer quels sont les liens qui unissent plusieurs images d'une même scène et comment utiliser ces liens. Indépendamment du problème géométrique, deux classes principales de solutions se dégagent, en fonction du fait que des connaissances *a priori* sur la scène observée soient disponibles ou non.

Construction de modèles

Si aucune connaissance sur la scène n'est disponible, alors la production d'un modèle de la scène consiste à construire complètement ce modèle. Passer des images bidimensionnelles à un modèle tridimensionnel nécessite alors au moins deux images, la projection 3D-2D entraînant en effet une perte d'informations. Les méthodes existantes dans ce cadre procèdent généralement suivant quatre étapes successives :

1) l'extraction de primitives dans les images, les primitives sont ici les projections dans les images des éléments constituant le modèle virtuel à construire. Ces primitives peuvent être des points, des segments ou des entités géométriques plus complexes (des courbes, *etc.*) ;

1. Voir le chapitre *visualisation et apparence.*

2) la mise en correspondance des primitives : il s'agit d'identifier les projections dans les différentes images d'un même élément de la scène ;

3) la triangulation : à partir des différentes projections images connues d'un élément et de connaissances sur les caméras (acquises et/ou connues *a priori*), il est possible de déterminer la position dans l'espace de cet élément par triangulation. C'est ici qu'intervient la géométrie pour déterminer les caractéristiques des projections ainsi que les positions respectives des différentes caméras impliquées dans le processus de modélisation ;

4) enfin lorsque l'on dispose de primitives dans l'espace, on peut alors construire un modèle virtuel regroupant l'ensemble de ces éléments et décrivant l'objet ou la scène observée. C'est la dernière étape de modélisation.

Ajustement de modèles

Lorsque des informations sur la scène sont disponibles, il est alors judicieux d'intégrer celles-ci dans le processus de modélisation. Ces informations peuvent être des connaissances partielles sur la scène telles que des angles ou des longueurs ; ou bien des connaissances plus complètes telle qu'un modèle *a priori* de la surface observée. La modélisation consiste alors à déformer le modèle tridimensionnel *a priori* de façon à ce que ses projections images correspondent aux images disponibles. Dans le cas de connaissances locales, le modèle *a priori* sera une entité intégrant ces connaissances, un parallélépipède par exemple pour modéliser des contraintes sur les angles et les longueurs. Nous verrons ainsi comment reconstruire des bâtiments à partir d'une seule image en ajustant des parallélépipèdes.

Les deux classes d'approches citées peuvent bien entendu être couplées pour intégrer le plus grand nombre d'informations possibles : des primitives extraites dans les images et des connaissances *a priori*, et permettre ainsi une modélisation plus riche.

La suite du document précise ces deux classes d'approches. L'objectif recherché n'est pas de faire une étude exhaustive des principes et approches existants mais plutôt de fournir une description intuitive du problème et des ses solutions à ce jour. Dans un premier temps, nous introduirons les outils géométriques utiles à la modélisation. Nous insisterons en particulier sur la géométrie d'une image et donc le calibrage d'une caméra, ainsi que sur la géométrie de plusieurs caméras et donc la détermination des positions respectives de plusieurs caméras. Ensuite, nous verrons comment construire un modèle et les différentes étapes nécessaires pour cela. Quelques approches reconnues seront présentées pour illustrer ce type de modélisation. Les techniques d'ajustement de modèles seront ensuite étudiées, aux travers notamment de quelques méthodes populaires. Nous conclurons enfin sur quelques perspectives dans le domaine de la modélisation à partir d'images.

1.2. Outils théoriques

Dans cette partie, nous décrivons le contexte géométrique nécessaire à la modélisation à partir d'images. Soulignons tout d'abord que la problématique de la modélisation d'une scène à partir d'images implique la modélisation des caméras ayant pris les images. La modélisation d'une scène requiert en effet « d'inverser » les projections ayant produit les images, soit de savoir de quelles manières ces images ont été formées et donc de connaître les caractéristiques des caméras impliquées.

Le premier paragraphe de cette section introduit le modèle géométrique et algébrique pour une caméra. La modélisation de scènes nécessite, dans le cas général, au moins deux images. Il est donc important de modéliser leurs relations spatiales, soit leurs positionnements relatifs au moment de l'acquisition. C'est, en effet, une étape nécessaire à la troisième étape du schéma de modélisation décrit dans l'introduction, à savoir la triangulation des primitives extraites des images. Le deuxième paragraphe de cette section décrit de quelle manière représenter le positionnement relatif de plusieurs caméras. Il s'agit de la *géométrie d'images multiples*. Le troisième paragraphe esquisse différentes manières d'estimer le positionnement relatif de caméras, à partir de différents types de données.

Citons ici les deux ouvrages [HAR 00, FAU 01] qui fournissent des informations plus détaillées sur les aspects géométriques de la modélisation à partir d'images. Tout au long de cette partie, nous nous intéressons à la description de projections de *points* qui sont les primitives les plus courantes et les plus *intuitives*. La généralisation à d'autres primitives peut, bien entendu, en être déduite.

1.2.1. *Géométrie d'une image*

Pour construire un modèle à partir d'images, il est nécessaire de savoir comment se forme une image, cela dans le but d'effectuer ensuite l'opération inverse et de reconstituer, à partir des images, la scène observée. Nous considérons ici les caméras numériques utilisant des capteurs CCD ou autres, une image étant alors un tableau de pixels, et nous précisons les modèles géométriques qui leurs sont applicables. Ces modèles restent valides dans le cas d'images analogiques (les appareils utilisant une pellicule photo), il suffit pour cela de numériser les images.

Les caméras sont composées de lentilles, de filtres, de circuits électroniques, *etc.* Un modèle complet nécessiterait la description de toutes ces composantes et de leurs contributions à la formation d'images. Nous nous focaliserons sur l'aspect géométrique uniquement, c'est-à-dire sur le fait de savoir comment un point de la scène est

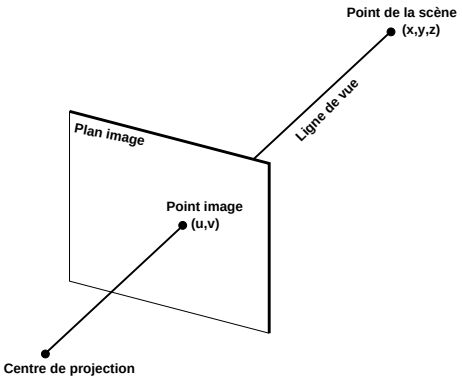

Figure 1.1. *Le modèle sténopé. Le plan image est ici situé entre le centre optique et la scène, mais le modèle pourrait être représenté de manière équivalente, au sens géométrique, avec le centre optique situé entre la scène et le plan image. L'image produite serait alors simplement à l'envers, c'est le cas par exemple de la rétine humaine.*

projeté sur un pixel dans l'image. Le modèle de caméra qui en découle permet de calculer, pour un point donné, à quel endroit dans l'image il sera visible[2]. À partir de ce modèle, on peut alors effectuer l'opération inverse, et essayer de déterminer, pour un point dans l'image, où se trouve le point de la scène correspondant. Bien entendu, ce que l'on peut obtenir à partir d'une seule image se résume à une demi-droite, *la ligne de vue*, sur laquelle doit se trouver le point (voir la figure 1.1). Plusieurs images, prises de plusieurs points de vue, sont alors nécessaires pour déterminer la position du point (cet aspect sera traité dans la partie 1.3.3 page 15).

Le modèle de caméra le plus utilisé actuellement est le modèle *sténopé* (appelé aussi modèle trou d'épingle). Dans ce modèle, une caméra est représentée par un *plan image* correspondant à la surface photosensible (le capteur CCD) et un *centre optique* ou *point focal* (voir la figure 1.1). Un point de la scène est projeté suivant la droite le reliant au centre optique, la ligne de vue, et le *point image* se trouve alors à l'intersection de cette droite avec le plan image.

Dans le but pratique d'effectuer des calculs numériques, une représentation algébrique est nécessaire. Les entités géométriques : points, droites et plans sont représentées pour cela à l'aide des coordonnées projectives, ou *coordonnées homogènes*. Dans cette représentation, un point d'un espace tridimensionnel est représenté par quatre

2. Ici, nous négligeons les effets d'une mauvaise mise au point ou du « flou », et nous supposons qu'un point se projette en un point exactement de l'image et non en une région de plusieurs pixels dans l'image.

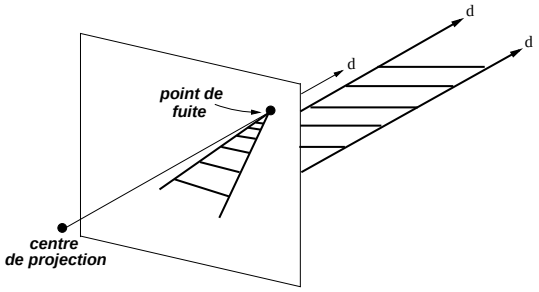

Figure 1.2. *Le point de fuite, projection de l'intersection à l'infini des rails parallèles, peut devenir visible dans le plan image d'une projection perspective. La géométrie projective permet de modéliser cet aspect et, plus généralement, les projections perspectives.*

coordonnées, qui sont définies à un facteur multiplicatif près. Les coordonnées homogènes constituent la clé de voûte de la géométrie projective et permettent notamment de représenter les entités à l'infini (ou à l'horizon) et donc de décrire les projections perspectives dans lesquelles ces entités à l'infini peuvent être visibles dans une image ; les rails parallèles du chemin de fer qui s'intersectent dans une image illustrent ce principe (voir la figure 1.2). L'opération de projection peut être représentée par une *matrice de projection* de dimension 3×4 [FAU 93]. Les coordonnées (u, v) de la projection image d'un point de l'espace (x, y, z) sont alors calculées par un simple produit matriciel :

$$\begin{pmatrix} w\,u \\ w\,v \\ w \end{pmatrix} = \begin{pmatrix} m_1 & m_2 & m_3 & m_4 \\ m_5 & m_6 & m_7 & m_8 \\ m_9 & m_{10} & m_{11} & m_{12} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

où w est un facteur multiplicatif.

La matrice de projection d'une caméra encapsule deux types d'informations distinctes. Tout d'abord, la projection dépend de la position et de l'orientation de la caméra par rapport à la scène, c'est ce que nous appellerons la *géométrie externe* de la caméra. Ensuite, la projection dépend d'autres propriétés indépendantes du positionnement : la distance focale ou la taille des pixels par exemple. Nous regroupons ces propriétés sous le terme *géométrie interne* de la caméra[3]. Les géométries interne et externe associées à une image permettent de calculer pour chaque point de cette image, la ligne de vue sur laquelle se trouve le point de la scène correspondant.

Le modèle sténopé est simple mais constitue, pour beaucoup d'applications, une approximation satisfaisante des caméras. Il existe des modèles plus simples encore,

3. On parle aussi de propriétés *extrinsèques* et *intrinsèques* des caméras.

ce sont les projections parallèles et en particulier orthographiques. Ces dernières sont parfois utilisées, mais elles restent cependant modélisables à partir du modèle sténopé générique. Pour une modélisation très précise de scènes ou pour certains types de caméras peu fréquentes (**fish eye** par exemple), des modèles plus complexes existent (voir [SLA 80] pour des exemples).

Quel que soit le modèle utilisé, le point important est qu'il permet de déterminer, pour un point d'une image, la ligne de vue correspondante sur laquelle se trouve le point origine de la scène.

1.2.2. *Géométrie de plusieurs images*

Nous savons qu'à partir d'une caméra et de son modèle, il est possible de déterminer, pour un point image, sa ligne de vue dans l'espace mais non la position exacte du point correspondant de la scène sur la ligne de vue. Pour déterminer cette position, il est nécessaire, dans le cas général, de considérer plusieurs images contenant le même point et de combiner les différentes lignes de vue correspondantes. Cela implique de connaître, en plus de la géométrie propre de chaque caméra, leurs positionnements relatifs les unes par rapport aux autres. Cette géométrie *commune* de différentes images est étroitement liée au problème de la correspondance : étant donnée la projection d'un point de la scène dans une image, que peut-on dire sur la position de ses projections dans d'autres images ? De fait, la connaissance de la géométrie commune contraint le problème de la correspondance, et réciproquement, la connaissance de correspondances entre images fournit des informations sur la géométrie commune. Nous reviendrons sur cet aspect dans la partie 1.3.2 traitant de la mise en correspondance de primitives.

1.2.2.1. *Géométrie épipolaire*

Nous considérons ici la géométrie de deux images. Soit p_1 un point dans la première image, qui est la projection d'un point P de la scène, dont la position est inconnue. Nous examinons ce qui peut être dit sur la position de la projection de P dans la deuxième image. Le modèle de la première caméra (et donc sa géométrie interne) nous donne la ligne de vue de p_1 (voir la figure 1.3-(a)). La seule information sur la position de P est le fait qu'il se trouve sur cette ligne de vue (voir la figure 1.3-(b)). Si nous projetons l'ensemble de la ligne sur la deuxième image (grâce à la connaissance (i) de sa géométrie interne, (ii) du positionnement relatif des caméras), nous obtenons alors toutes les positions possibles du point recherché. Si le modèle de caméra est le modèle sténopé (voir le paragraphe précédent), les projections des points de la ligne de vue constituent une droite d_2 dans la deuxième image (voir la figure 1.3 (b)). La recherche de la correspondance de p_1 peut donc se limiter à la droite d_2 et devient beaucoup moins complexe qu'une recherche exhaustive dans toute l'image.

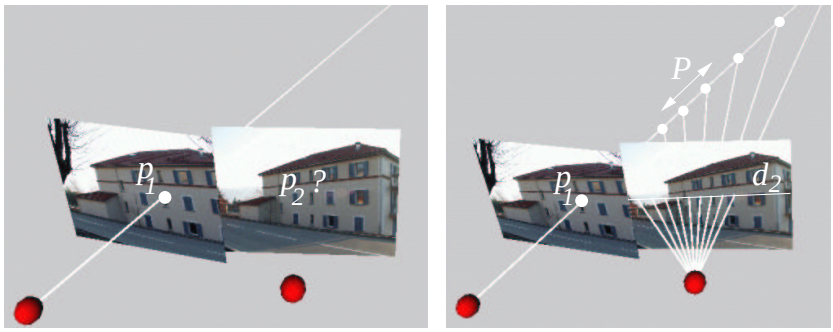

Figure 1.3. *La géométrie épipolaire caractérise le fait que les correspondants potentiels dans l'image* 2 *d'un point* p_1 *de l'image* 1 *sont situés sur une droite* d_2 *appelée droite épipolaire.*

Cette relation géométrique entre des points image correspondant dans deux images, est appelée la *géométrie épipolaire*, et la droite d_2 la *droite épipolaire* de p_1 dans la deuxième image. La construction de la droite épipolaire présentée est basée sur une reconstruction explicite de la ligne de vue dans l'espace, et sur sa projection dans la deuxième image. Ces opérations peuvent être regroupées sous la forme d'une seule relation projective 2D-2D, qui transforme les vecteurs de coordonnées de points dans la première image, en vecteurs de coordonnées de lignes épipolaires dans la seconde. Elle peut être représentée par une matrice de dimension 3×3, appelée *matrice fondamentale* [LUO 96]. La connaissance de cette matrice associée à celle de la géométrie interne des caméras est équivalente à la connaissance du déplacement entre les caméras, et donc de leurs positions relatives.

Si plus de deux images sont considérées, des relations plus complexes que la géométrie épipolaire existent. Par exemple, la géométrie de trois images peut être décrite par des tenseurs de dimension $3 \times 3 \times 3$, souvent appelés *tenseurs tri-focaux* [SHA 95, TRI 95]. Ils permettent de restreindre encore plus l'espace de recherche lors de la mise en correspondance.

En résumé, la connaissance de la géométrie des caméras est utile pour la recherche de correspondances entre des images. Réciproquement, la connaissance de correspondances donne des indices sur la géométrie des caméras. Il est important de noter que des correspondances entre images suffisent pour estimer cette géométrie – aucune connaissance de la scène n'est nécessaire pour cela. Cet aspect trouve des applications pratiques : par exemple estimer les positions respectives de deux caméras dont les caractéristiques intrinsèques sont connues à partir de quelques points images mis en correspondance [HAR 95, FAU 93].

Figure 1.4. *Un objet de calibrage très largement utilisé : le damier. Les sommets constituent en effet des primitives qui sont facilement détectables dans les images, et dont les positions dans la scène sont connues.*

1.2.3. *Estimer la géométrie des caméras*

Nous passons brièvement en revue les principaux moyens d'estimer la géométrie des caméras, dans le contexte de la modélisation de scènes.

Le moyen le plus simple, le *calibrage*, consiste à utiliser un objet dont la géométrie est connue. A partir d'une seule image de cet objet, la géométrie complète de la caméra peut alors être estimée : sa géométrie interne et son positionnement relatif par rapport à l'objet [FAU 87, TSA 87, STU 99]. Si la modélisation de la scène est effectuée à l'aide de plusieurs caméras *statiques*, la géométrie complète de ce système peut être estimée de la même manière en plaçant un objet connu dans le champ de vue commun des caméras. Des logiciels du domaine public[4] permettent, en particulier, de calibrer à l'aide d'une grille (voir la figure 1.4). Les caméras peuvent donc être calibrées et leurs positions relatives peuvent être déduites des positions relatives à l'objet de calibrage.

Ce scénario n'est pas toujours applicable, en particulier si, au cours de l'acquisition d'images, les caméras se déplacent, ou si le zoom ou la mise au point sont modifiés (et, par voie de conséquence, la géométrie interne), ou bien si l'on souhaite éviter le processus de calibrage pour des raisons de temps ou d'équipement. Dans ce cas, les outils de la géométrie multi-images doivent être mis en œuvre. En effet, comme nous l'avons vu dans la partie précédente, des correspondances entre images, peuvent permettre d'estimer la géométrie des caméras, même si la structure de la scène n'est pas connue. Dans le cas le plus général, les géométries interne *et* externe des caméras doivent être estimées à partir de ces correspondances. On parle alors d'*auto-calibrage*.

4. `http ://intel.com/research/mrl/research/opencv/`

Il a été prouvé, au début des années 90, que l'auto-calibrage était effectivement possible [FAU 92, HAR 92]. Plusieurs approches ont été proposées depuis ([FUS 00] les détaille). La mise en œuvre pratique de l'auto-calibrage reste néanmoins une opération délicate.

Une solution intermédiaire consiste à pré-calibrer les caractéristiques intrinsèques des caméras (la géométrie interne) que l'on suppose fixes, et à déterminer ensuite la géométrie externe uniquement. Dans ce cas, un auto-calibrage complet n'est pas nécessaire, la géométrie interne des caméras étant connue. On parle alors d'*estimation du mouvement*. Il s'agit d'un problème plus simple, mais qui interdit l'usage du zoom (car il modifie la géométrie interne) et qui requiert toujours une étape préliminaire de calibrage. Le choix de l'approche pour l'estimation de la géométrie des caméras, dépend du contexte de l'application envisagée et des contraintes qui en découlent : temps, précision, nombre de caméras, *etc.* Le tableau 1.1 récapitule les principales opérations géométriques existantes, et ce qui doit être connu et ce qui est déterminé pour chaque opération.

Opération	Géométrie de la scène	Géométrie externe des caméras	Géométrie interne des caméras
Calibrage	connue	déterminée	déterminée
Calcul de pose	connue	déterminée	connue
Triangulation	déterminée	connue	connue
Estimation du mouvement	inconnue	déterminée	connue
Reconstruction non-métrique	déterminée	inconnue	inconnue
Auto-calibrage	inconnue	inconnue	déterminée

Tableau 1.1. *Les différentes opérations géométriques et les contextes correspondants.*

1.3. Construction de modèles

La construction de modèles repose sur des informations extraites dans les images uniquement, sans connaissances *a priori* sur la scène observée. Nous supposons par contre ici que la géométrie des caméras est connue (voir la partie précédente). Comme cela a été dit en introduction, la construction de modèles suit, en général, quatre étapes principales qui sont : l'extraction de primitives dans les images, la mise en correspondance des primitives extraites d'une image à une autre, la triangulation ou la reconstruction de ces éléments correspondants dans l'espace (c'est-à-dire les éléments dans l'espace ayant pour projections dans les images les primitives extraites), et enfin la construction d'une surface/modèle à partir des éléments reconstruits. Ce schéma est

celui de la plupart des approches de construction, même si, comme nous le verrons dans la partie applications, quelques approches ne le vérifient pas scrupuleusement (les approches basées sur les contours occultants en particulier).

À ces quatre étapes s'ajoute un *pré-traitement* des images. L'information disponible dans les images est en effet loin d'être parfaite pour diverses raisons : numérisation de l'information, compression éventuelle, déformation des objectifs, *etc*.. Il est donc nécessaire de pré-traiter les images pour en améliorer la qualité en vue d'extraire des primitives. Le pré-traitement d'images constitue un domaine à part entière dont les applications dépassent le cadre de la construction de modèles. Son étude sort donc du contexte de ce document mais le lecteur intéressé pourra consulter les ouvrages suivants pour obtenir plus d'informations [PRA 78, COC 95].

Dans les parties suivantes nous détaillons les étapes mentionnées et présentons différentes applications reconnues de construction de modèles.

1.3.1. *Extraction de primitives*

L'extraction de primitives est la première étape de la construction d'un modèle à partir d'images. Ces primitives peuvent être des points, des segments ou des courbes et peuvent correspondre à des cibles placées artificiellement dans la scène, ou à des points d'intérêt dans les images ainsi qu'à des contours naturellement présents dans les images. Leur extraction d'images peut se faire de façon manuelle ou automatique, ou par soustraction du fond pour un contour en mouvement. Les parties suivantes explicitent les différentes approches d'extraction automatique de primitives.

1.3.1.1. *Extraction de contours*

Les contours constituent des indices riches, au même titre que les points d'intérêt, pour toute interprétation du contenu d'une image. Les contours dans une image proviennent des discontinuités de la fonction de réflectance (texture, ombre), et des discontinuités de profondeur (les bords de l'objet). Ils sont caractérisés par des discontinuités de la fonction d'intensité dans les images. Les méthodes existantes se focalisent principalement sur les discontinuités d'ordre 0 (voir la figure 1.5). Le principe de la détection de contours repose donc sur l'étude des dérivées de la fonction d'intensité dans l'image et la recherche des extréma locaux du gradient de la fonction d'intensité et des passages par zéro du laplacien (filtres de Prewitt, Sobel, Marr-Hildreth, Canny, Deriche [COC 95]).

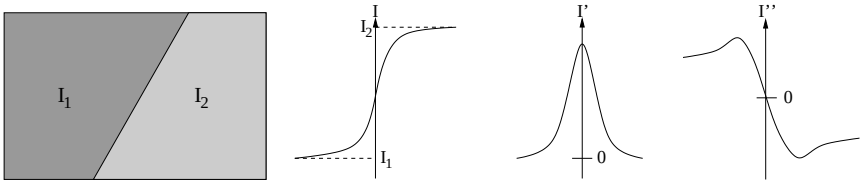

Figure 1.5. *Un contour en forme de marche, la fonction d'intensité au voisinage de ce contour ainsi que ses dérivées première et seconde.*

1.3.1.2. *Extraction de points d'intérêt*

Les points d'intérêt, dans une image, correspondent à des doubles discontinuités de la fonction d'intensité. Celles-ci peuvent être provoquées, comme pour les contours, par des discontinuités de la fonction de réflectance ou des discontinuités de profondeur. Ce sont par exemple : les coins, les jonctions en T ou les points de fortes variations de texture.

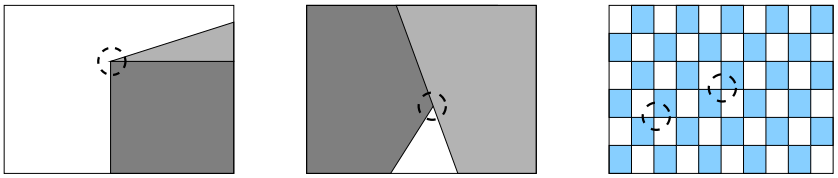

Figure 1.6. *Différents types de points d'intérêt : coins, jonctions en T et points de fortes variations de texture.*

Les points d'intérêt présentent quelques avantages par rapport aux contours, ils sont en particulier présents dans un plus grand nombre d'images. Leurs détections consiste à rechercher les doubles discontinuités de la fonction d'intensité (le détecteur de Harris [HAR 88, SCH 98] par exemple).

1.3.1.3. *Soustraction de fond*

Dans le cas d'une séquence d'images : une vidéo ou des images rapprochées, prises à partir d'une caméra fixe, il est possible d'extraire un contour en mouvement. L'idée est qu'un pixel du fond de l'image a une intensité qui varie peu dans le temps si la caméra est fixe, alors qu'un pixel correspondant à un objet en mouvement a une intensité qui varie fortement. La détection de contours en mouvement va donc consister à vérifier, pour un pixel, les variations de sa fonction d'intensité. Plusieurs approches existent, de la solution *naïve* qui consiste à soustraire l'image précédente à l'image courante et à appliquer un seuillage, à la solution plus complexe faisant intervenir un

Figure 1.7. *Une image, les contours extraits (filtre de Canny) et les points d'intérêt extraits (détecteur de Harris).*

mélange de Gaussiennes pour modéliser la fonction d'intensité, ou bien faisant intervenir un filtrage des intensités par une prédiction linéaire (filtre de Wiener) (voir [TOY 99] pour une étude comparative). La figure 1.8 illustre ce principe.

Figure 1.8. *un exemple de soustraction de fond d'image basé sur un seuillage de la variation chromatique des pixels [HOR 99].*

1.3.2. *Mise en correspondance*

Lorsque des primitives ont été extraites dans plusieurs images d'une scène, il est alors nécessaire de les *mettre en correspondance*, c'est-à-dire d'identifier dans les différentes images les primitives correspondant à une même entité de la scène. C'est une étape fondamentale de la modélisation à partir d'images, qui est parfois réalisée manuellement. Nous nous intéressons, dans cette partie, aux approches automatiques de mise en correspondance s'appliquant à des primitives de type point. (On pourra consulter [SCH 00] pour une méthode automatique s'appliquant à des primitives de type segment de droite ou courbe).

1.3.2.1. *Utilisation de cibles codées*

Il s'agit typiquement de cibles circulaires et souvent auto-réfléchissantes, qui sont étiquetées par un code barre circulaire (voir la figure 1.9). Ce code peut être déchiffré automatiquement par un traitement du signal et permet une mise en correspondance aisée entre des images. Ces cibles sont fréquemment utilisées en photogrammétrie où la précision et la fiabilité sont essentielles et dans des applications où le déploiement de cibles sur des objets est réalisable.

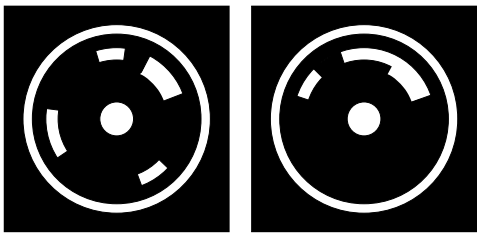

Figure 1.9. *Exemples de cibles codées : le cercle extérieur sert à détecter une cible dans l'image et le code barre à l'identifier. Le cercle intérieur (la position de son centre) est la primitive qui est finalement utilisée pour les calculs géométriques.*

1.3.2.2. *Mise en correspondance automatique*

C'est l'un des problèmes ayant suscité le plus de recherches en vision par ordinateur. Il se résume essentiellement à la tâche suivante : identifier, dans une image que nous appellerons *image* 2, le pixel p_2 qui correspond au même point de la scène qu'un pixel donné p_1 dans ce que nous appellerons l'*image* 1. On parle d'une mise en correspondance *dense* si les correspondants de tous les pixels d'une image doivent être trouvés. La majorité des approches de modélisation ne nécessite qu'une mise en correspondance *éparse*, c'est-à-dire entre quelques points de l'image, des points d'intérêt par exemple (voir la partie 1.3.1.2 page 12).

Les méthodes de mise en correspondance sont nombreuses, nous évoquons seulement les principes communs à la plupart d'entre elles. La décision de savoir si deux

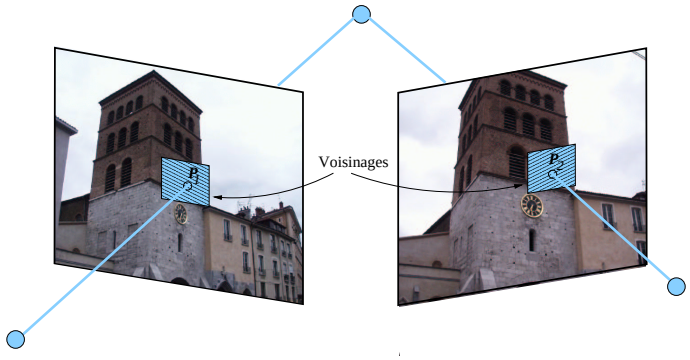

Figure 1.10. *Deux pixels p_1 et p_2 sont considérés comme correspondants si leurs voisinages sont similaires, c.à.d si les pixels des deux voisinages ont des niveaux d'intensité similaires. Ces voisinages peuvent être de simples zones rectangulaires ou bien des quadrilatères plus généraux prenant en compte une déformation affine ou projective des voisinages.*

pixels se correspondent est généralement basée sur la comparaison des fonctions d'intensité au voisinage des pixels considérés (voir la figure 1.10). Les approches diffèrent ensuite sur les voisinages considérés : de la simple fenêtre rectangulaire autour du pixel, au voisinage *corrigé* par transformations affines ou projectives pour la prise en compte des déformations perspectives. Un autre aspect important est la robustesse de la mise en correspondance aux changements d'illumination apparents, par exemple dans le cas d'images prises avec des caméras possédant des gains différents. Les approches actuelles normalisent, pour cela, les intensités sur les voisinages considérés. Pour trouver le correspondant d'un pixel de l'image 1, il faut, *a priori*, parcourir l'ensemble de l'image 2 à la recherche du pixel qui minimise la différence d'intensité appliquée au voisinage. Si par contre, le positionnement relatif des deux caméras est connu, on peut réduire l'espace de recherche de manière importante, en utilisant la géométrie épipolaire (voir la partie 1.2.2.1 page 7). Nous savons en effet que le correspondant de p_1 se situe nécessairement sur la droite épipolaire de p_1 dans l'image 2. Il est à noter que l'utilisation de la géométrie épipolaire réduit non seulement le temps de calcul de la mise en correspondance, mais aussi la probabilité d'erreurs.

1.3.3. *Triangulation*

Par triangulation, nous entendons ici le calcul de la position 3D d'une primitive, à partir de ses projections dans deux, ou plus, images (obtenues par extraction et mise en correspondance, voir les parties 1.3.1 et 1.3.2 page 11 et 14) et de la connaissance de la géométrie des caméras (voir la partie 1.2.3 page 9). Le principe de la triangulation s'illustre très simplement dans le cas des points : la géométrie des caméras détermine, pour chaque image, la ligne de vue du point image considéré. La position du point dans la scène est alors à l'intersection des lignes de vue (voir la figure 1.11).

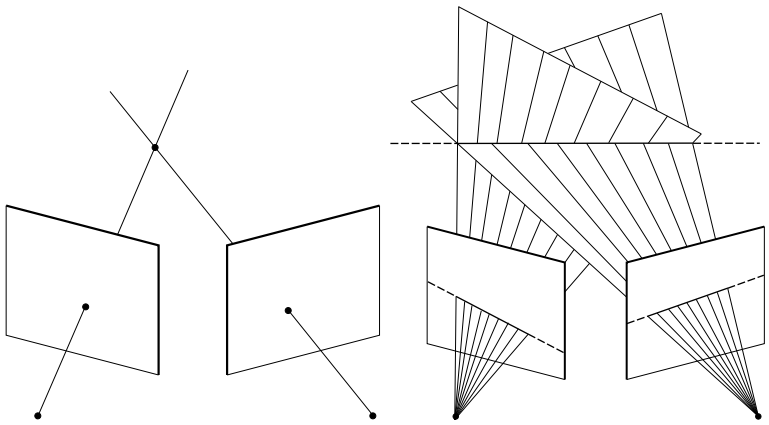

Figure 1.11. *Triangulation d'un point ou d'une droite.*

Bien entendu, en raison du bruit présent dans les données, des imprécisions de l'extraction de points ou de la géométrie des caméras par exemple, il n'y pas d'intersection exacte. La solution consiste alors à chercher le point qui minimise un critère d'optimalité prenant en compte la distance aux lignes de vue et, éventuellement, un modèle probabiliste du bruit (voir [HAR 97]).

La triangulation de droites, ou segments de droites, s'effectue de manière similaire. On considère alors non plus la ligne de vue associée à un point, mais le plan de vue associé à une droite. Les plans de vue définis par deux, ou plus, images d'une droite, permettent de déterminer la position de cette droite par intersection (voir la figure 1.11). La triangulation de primitives plus complexes, telles des courbes, peut en principe être effectuée de manière analogue, mais des solutions analytiques sont plus difficiles à obtenir.

Notre schéma général sépare la modélisation de la scène de celle des caméras, qui en effet est supposée être effectuée au préalable. Ceci ne permet pas d'exploiter toutes les redondances existant entre les paramètres (caméras et scène). Pour améliorer la précision, on effectue souvent une étape d'optimisation à la suite de la triangulation. Cette étape consiste à optimiser simultanément tous les paramètres, des modèles de la scène ainsi que des caméras impliquées. Ce processus est appelé *ajustement de faisceaux* [TRI 99]. Il est particulièrement utile dans le cas d'une estimation initiale peu précise de la géométrie des caméras, obtenue par auto-calibrage ou par estimation du mouvement par exemple.

1.3.4. *Construction de surfaces*

Les différentes étapes présentées dans les parties précédentes permettent de reconstruire dans l'espace des primitives appartenant à la scène observée. Comme nous l'avons vu, ces primitives peuvent être des points, des segments ou bien des courbes. Il s'agit ensuite de produire un modèle à partir de cet ensemble de primitives dans l'espace, c'est-à-dire d'obtenir une description de la scène sous la forme d'une surface. Cette description permet en particulier de plaquer des textures et d'obtenir ainsi un modèle photo-réaliste. Plusieurs approches sont possibles : de l'approximation de l'ensemble des points par une surface à leur interpolation par un modèle constitué de facettes polygonales.

1.3.4.1. *Par approximation*

L'approximation consiste à rechercher une surface qui, sans contenir nécessairement les points 3D du modèle, passe à proximité de ces points. Il est possible pour cela de *déformer* un modèle de surface connu *a priori*. Néanmoins, ce type d'approche présente peu d'intérêt dans le contexte qui est le notre, puisque les données qui nous concernent sont, à l'origine, des primitives images et que la déformation du modèle *a priori* devrait se faire non pas en fonction de données 3D calculées mais en fonction des données 2D originales. C'est d'ailleurs l'objet d'une partie ultérieure de ce chapitre (voir la partie 1.4 page 24). Un autre type d'approche par approximation consiste à utiliser les points 3D du modèle pour estimer une fonction de distance à la surface, le noyau de cette fonction étant ensuite polygonalisé (à l'aide de l'algorithme des marching cubes) pour produire un modèle formé de facettes polygonales [HOP 92, CUR 96].

1.3.4.2. *Par interpolation*

L'interpolation des primitives reconstruites est une méthode très largement utilisée. Elle consiste généralement à *trianguler* les primitives, soit obtenir un modèle représenté par une collection de facettes triangulaires. La triangulation des primitives peut, bien entendu, s'effectuer manuellement. Cette opération s'avère néanmoins fastidieuse si les primitives sont nombreuses. Il existe des approches, dans le domaine de la géométrie algorithmique, qui permettent de trianguler automatiquement des points dans l'espace et d'obtenir une surface [AME 98, BOY 00]. Ces approches ont un domaine d'application plus large que la vision, mais en revanche, elles ne prennent pas en compte l'information contenue dans les images, à savoir l'information photométrique. Des approches plus spécifiques au domaine de la vision proposent de rechercher une triangulation optimale au sens de la *cohérence image*. Les éléments de cette triangulation, les triangles, doivent être tels que leurs différentes projections images soient cohérentes, ou en d'autres termes contiennent des informations photométriques similaires [MOR 00] (voir la figure 1.12).

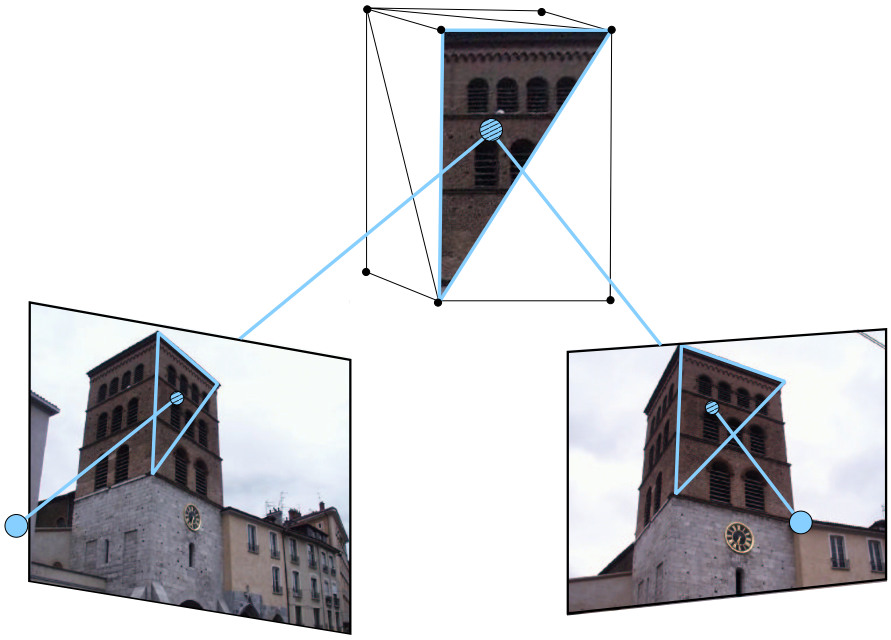

Figure 1.12. *La cohérence image du modèle signifie que les projections d'une facette décrivent la même partie du modèle dans les images. Ici le contenu de la projection de la facette ~~tri~~angulaire du modèle est le même dans les deux images. Cette cohérence est recherchée lors de la détermination des facettes constituant le modèle.*

1.3.4.3. *Placage de textures*

Mentionnons ici le placage de textures qui est un aspect important de la modélisation à partir d'images puisqu'il confère aux modèles un réalisme important : le photo-réalisme. Le principe est simple, prenons un élément du modèle : une facette triangulaire pour simplifier, et considérons sa projection dans une image. Le placage de texture consiste alors à effectuer l'opération inverse, c'est-à-dire à projeter le contenu de la facette dans l'image sur le modèle tridimensionnel (voir la figure 1.12 par exemple). Cette opération ne nécessite pas toujours la connaissance exacte de la relation de projection existant entre le modèle tridimensionnel et l'image. Deux cas de figure peuvent en effet se présenter : (i) la projection image est une projection orthographique ou parallèle, la transformation entre la facette du modèle et son image projetée est dans ce cas une transformation affine. Le placage de texture s'effectue alors par une relation affine entre la facette et le modèle, sans connaissances autres que le modèle tridimensionnel, l'image et les positions dans l'image des sommets projetés du modèle. La plupart des outils de visualisation existants réalisent cela en temps réel (les visualiseurs VRML par exemple) ; (ii) la projection est perspective.

Dans ce cas, la relation entre la facette image et celle du modèle n'est plus affine et le placage de texture par les outils standards nécessite, en général, une *rectification affine* [HAR 00], c'est-à-dire une transformation de la facette image telle que la relation image-modèle devienne affine. Cette rectification affine nécessite la connaissance de caractéristiques perspectives de l'image, par exemple les points de fuites associés à la facette de texture lorsque celle-ci est rectangulaire.

A noter ici que la cohérence image, dont il a été question précédemment pour déterminer les facettes du modèle, se traduit simplement par le fait que les textures, dans plusieurs images, d'une même facette du modèle doivent être similaires.

1.3.5. *Quelques approches de construction de modèles*

Nous présentons dans cette partie quelques approches de construction de modèles qui sont largement utilisées par la communauté de recherche en vision et dans des logiciels commerciaux. Ces approches peuvent être spécifiques à un contexte particulier : des objets de dimensions petites ou moyennes dans le cas des contours occultants par exemple, l'adjonction d'éléments tel qu'un éclairage spécifique dans le cas de lumières structurées. Nous précisons à chaque fois le contexte d'application de la méthode présentée.

1.3.5.1. *À partir de cibles*

Le but traditionnel de la photogrammétrie est d'obtenir des mesures à partir de cibles disposées dans la scène et clairement identifiables, la création de surfaces n'est en général pas envisagée. Des cibles, comme par exemple celles décrites dans la partie 1.3.2.1 (page 14), sont donc déployées à différents points. Leurs extractions et mises en correspondance s'effectuent simplement : à l'aide de contrastes très forts dans l'image et à l'utilisation d'un code barre par exemple. La géométrie des caméras est connue de manière approximative est peut être raffinée à l'aide des cibles par un ajustement de faisceaux (voir la partie 1.3.3 page 15). Le contexte d'application de ces approches est très large, la limitation ne réside pas, en effet, dans la structure de la scène mais dans la présence nécessaire de cibles identifiables dans la scène.

1.3.5.2. *À partir de segments ou de points*

De nombreuses approches existent dans ce cadre et consistent à extraire et mettre en correspondance dans plusieurs images des primitives qui sont des segments de droite (voir la figure 1.13), ou bien des points d'intérêt. Dans ce dernier cas, il existe des approches qui permettent de calibrer (voir la partie 1.2.3 page 9) et de reconstruire uniquement à partir des mises en correspondance de points [POL 98].

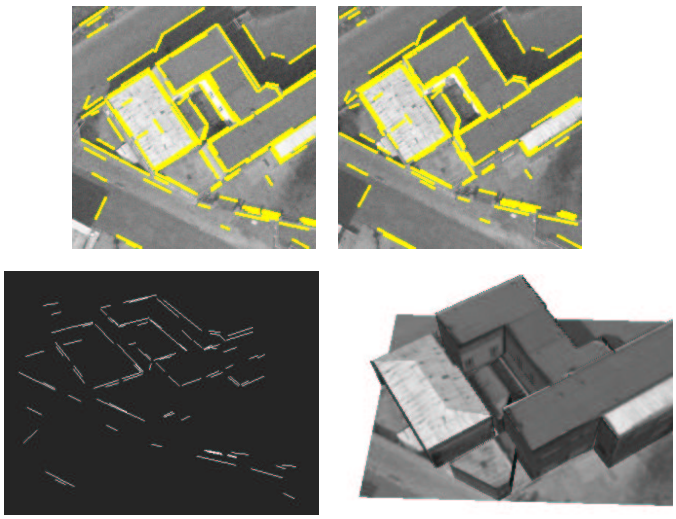

Figure 1.13. *Un exemple de construction de modèles à partir d'images aériennes : en haut le détail de deux des images de la séquence utilisée montrant les segments de droites extraits, dessous à gauche les segments de droites reconstruits et dessous à droite le modèle texturé correspondant (images de C. Schmid et A. Zisserman [SCH 00]).*

Citons aussi plusieurs systèmes commerciaux dont PhotoModeler[5] ou ImageModeler[6] qui permettent la création de modèles sans utilisation de cibles, mais au prix d'une interaction humaine relativement lourde : l'utilisateur doit manuellement extraire et mettre en correspondance toutes les primitives qui doivent composer le modèle. La modélisation des caméras et la triangulation se font automatiquement. Il est ensuite possible de créer des surfaces de différents types (des splines, des facettes planes, *etc.*), mais là encore l'intervention de l'utilisateur est nécessaire. Une extension évidente de ces systèmes concerne l'automatisation de l'extraction et la mise en correspondance. Plusieurs laboratoires de recherche ont développé des systèmes performants qui ont débouché sur des produits commerciaux (boujou[7] ou MatchMover[8]). La création automatique de surfaces reste néanmoins un problème difficile. Le principal domaine d'application de ces systèmes commerciaux concerne les effets spéciaux pour la post-production de films, où l'objectif est le mélange du réel et du virtuel.

5. http ://www.photomodeler.com
6. http ://www.realviz.com
7. http ://www.2d3.com
8. http ://www.realviz.com

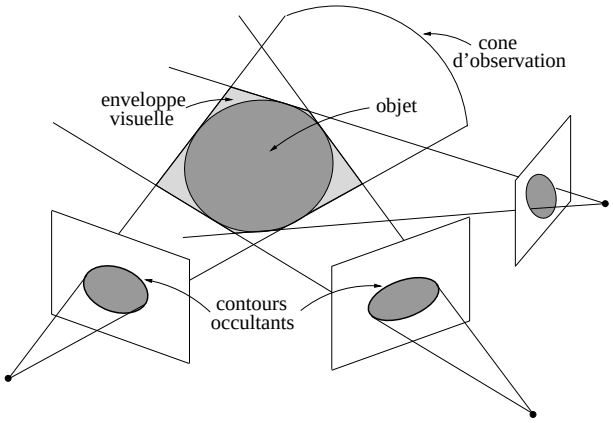

Figure 1.14. *Les cônes d'observation d'un objet et l'enveloppe visuelle intersection de ces cônes dans l'espace.*

Ces méthodes s'appliquent dans un contexte très large puisque la seule contrainte est d'avoir des primitives présentes dans plusieurs images. Le calibrage n'est, en particulier, pas toujours nécessaire.

1.3.5.3. *À partir des contours occultants*

Parmi les approches populaires, citons les approches utilisant les contours occultants. Un contour occultant correspond à la *silhouette* d'un objet dans une image (voir la figure 1.14). En d'autres termes, le contour occultant d'un objet délimite dans l'image la partie visible de la surface de l'objet. Il peut être obtenu par soustraction de fond (voir la partie 1.3.1.3 page 12) par exemple. Dans le cas d'un objet possédant une surface lisse, ces contours ne sont pas fixes sur la surface mais se déplacent en fonction du point de vue. Une difficulté est ici que l'on ne puisse pas effectuer de triangulation comme cela a été vu précédemment puisque le contour observé est différent pour chaque image. Le schéma de construction de modèles vu précédemment ne fonctionne donc pas ici. Par contre, puisque le contour occultant délimite la partie visible de la surface de l'objet, il est possible de construire une approximation de cette surface à partir de ces contours. Considérons pour cela un contour dans une image, ce contour délimite dans l'espace un volume dans lequel se trouve l'objet observé, on parle de *cône d'observation*. Si plusieurs images sont disponibles, alors l'intersection des différents cônes d'observation correspondants constitue une approximation de l'objet observé (voir la figure 1.14). En fonction des points de vue (nombre et positions) cette approximation, on parle d'*enveloppe visuelle* [LAU 94], est plus ou moins précise. Il existe plusieurs approches pour déterminer cette enveloppe visuelle :

1) Approches volumiques : l'idée ici est de *sculpter* un volume dans l'espace. A partir d'un volume initial composé de cellules élémentaires, les *voxels*, on élimine les

cellules se projetant, dans une, ou plus, des différentes images considérées, à l'extérieur de la zone délimitée par le contour occultant [SZE 93, NIE 94] (voir la figure 1.15 par exemple).

Figure 1.15. *Un exemple de construction par sculpture de volume dans l'espace. À gauche, une des images utilisées (l'objet est sur une table tournante), au milieu le modèle VRML produit et à droite le même modèle texturé (images de G. Cross et A. Fitzgibbon [FIT 98]).*

2) Approches surfaces : l'idée, cette fois-ci, est de déterminer l'enveloppe visuelle correspondant aux différents points de vue et délimitant le volume intersection des cônes d'observations. Ces approches peuvent être basées sur des algorithmes de calcul d'intersection de polyèdres [SUL 98] ou de polygones [MAT 01]. Une autre direction consiste à déterminer la topologie de l'enveloppe visuelle, c'est-à-dire les sommets et les segments de courbes reliant ces sommets [LAZ 01]. Cette topologie est intéressante à plusieurs titres : (i) elle est indépendante de toutes considérations de précision ; (ii) elle est indépendante de la géométrie interne des caméras et peut donc être déterminée sans connaissance sur cette dernière.

Ces approches sont bien adaptées aux objets de petites et moyennes dimensions pour lesquels il est facile soit de déplacer une caméra autour de l'objet soit de déplacer l'objet lui-même. Elles nécessitent, pour la plupart, le calibrage complet des caméras (géométrie interne et externe).

1.3.5.4. *À partir de l'information photo-métrique*

Une approche récente de modélisation : le *space carving* [KUT 00b, KUT 00a], utilise, comme précédemment, une discrétisation de l'espace en cellules, les voxels, et une technique de sculpture de ce volume par élimination des cellules. La différence réside dans le critère mis en œuvre pour éliminer les cellules. Il repose ici sur la cohérence image entre les projections d'un même voxel. L'idée est qu'un voxel qui appartient à l'objet observé doit se projeter, dans les images où il est visible, en des pixels dont les valeurs photométriques sont cohérentes (égales lorsqu'il s'agit d'une surface Lambertienne). Si un voxel ne vérifie pas ce critère, il est alors éliminé (voir

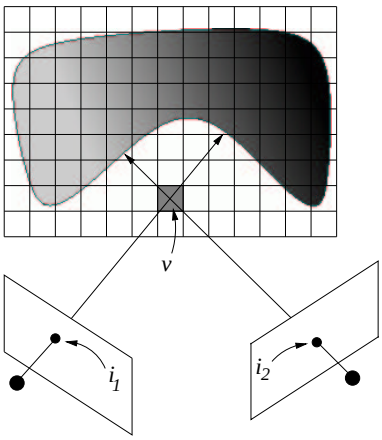

Figure 1.16. *Principe du space carving : un voxel v est éliminé lorsque ses projections, dans les images où il est visible, ne correspondent pas au même élément de l'objet et présentent donc des valeurs photométriques i_1 et i_2 incohérentes.*

la figure 1.16). Ce type d'approches produit des modèles constitués de voxels colorés. Un intérêt par rapport aux techniques basées sur les contours occultants (voir la partie 1.3.5.3 précédente) est que les parties concaves de l'objet sont traitées alors qu'elles n'apparaissent pas dans une enveloppe visuelle. Une application consiste donc à utiliser un volume initial pré-sculpté à l'aide de contours extraits (une enveloppe visuelle), et à améliorer ce modèle en affinant le volume sculpté à l'aide du space carving.

Ces approches ont un contexte d'applications assez large, elles nécessitent par contre la connaissance de la géométrie des caméras.

1.3.5.5. *À l'aide de lumières structurées*

Une approche flexible de construction de modèles consiste à utiliser un système d'illumination projetant des motifs sur la scène à modéliser. Ces motifs servent alors de cibles ou de points d'intérêt qui sont facilement détectables et identifiables dans les images. Les motifs utilisés sont variés, il peut s'agir de lignes ou d'une grille, projetés à l'aide d'un laser. L'approche se décompose ensuite suivant les étapes mentionnées précédemment, à savoir : l'extraction des cibles (les sommets de la grille par exemple) dans les images, leurs mises en correspondance ou identifications dans les différentes images, leurs reconstructions dans l'espace et enfin la construction d'un modèle à partir de ces cibles. Il est en particulier possible de reconstruire des parties du corps humain de cette manière dont le visage. De nombreux produits commerciaux basés sur ce principe existent. Ce sont notamment les *scanners 3D* qui utilisent des lasers et produisent éventuellement des modèles texturés.

Ces approches sont bien adaptées aux objets de petites dimensions sur lesquels il sera facile de projeter des motifs. Elles nécessitent un calibrage complet des caméras (géométrie interne et externe) et éventuellement du système caméra-lumière structurée.

1.4. Ajustement de modèles

À la différence des approches de construction de modèles, les approches d'ajustement de modèles reposent sur la connaissance d'informations *a priori* sur le modèle. Dans beaucoup d'applications le contexte est, en effet, connu de manière assez précise : les objets considérés appartiennent par exemple à une classe de modèles (des bâtiments par exemple), ou bien des contraintes existent et sont facilement exploitables : la coplanarité de plusieurs primitives ou la connaissance des angles entre plusieurs primitives par exemple. L'idée est alors d'utiliser ces connaissances dans le processus de modélisation, et d'éviter ainsi une ou plusieurs des étapes nécessaires à la construction de modèles telles que : l'estimation de la géométrie des caméras, la mise en correspondance ou la construction de surfaces. L'intérêt est non seulement de simplifier le processus de modélisation mais aussi de le rendre plus robuste.

Dans cette partie, nous décrivons tout d'abord brièvement le principe de l'ajustement de modèles puis nous illustrons ce principe par l'étude de différentes approches d'ajustement de modèles. Nous ne cherchons pas ici à décrire l'ensemble des approches existantes mais plutôt à fournir quelques idées directrices au travers de quelques applications exemplaires.

1.4.1. *Principe*

Le principe de l'ajustement de modèles est donc de partir d'un modèle tridimensionnel *a priori* d'un objet de la scène ou de la scène elle-même, et de déformer ce modèle de façon à ce que sa projection dans les images considérées corresponde aux informations disponibles dans les images.

1.4.1.1. *Identification dans les images*

Un premier aspect du problème concerne ces informations images. Il est en effet nécessaire d'identifier, dans chaque image considérée, la projection du modèle, ou d'une partie du modèle si des occultations du modèle se présentent. Cette identification peut se faire à la main. Dans ce cas, l'utilisateur fournit les positions du modèle (ou d'une partie du modèle), dans les images. Il n'y a donc pas ici d'extraction automatique de primitives dans les images, ces dernières sont fournies par l'utilisateur. Une autre approche consiste à identifier de manière automatique les projections du modèle dans les images. On peut, pour cela, effectuer une extraction automatique de primitives (points ou contours, partie 1.3.1 page 11) et identifier parmi les primitives

extraites celles correspondant au modèle. Bien entendu, cette deuxième approche reste délicate à mettre en œuvre dans un contexte pratique et ne concerne donc que quelques méthodes.

1.4.1.2. *Ajustement*

Ensuite, lorsque les projections du modèle ont été identifiées dans les images, il s'agit d'ajuster le modèle, soit de modifier les paramètres le définissant, jusqu'à ce que ses projections correspondent à celles identifiées. Cette étape d'ajustement est une étape d'*optimisation* où l'on va chercher à minimiser un critère qui est fonction des paramètres du problème, et qui prend en compte l'adéquation du modèle aux données, à savoir les projections images du modèle. Un point fondamental ici est que ce processus d'optimisation fait intervenir non seulement les paramètres du modèle, mais aussi ceux des caméras puisque les projections du modèle varient en fonction des caractéristiques des caméras. Cette étape d'ajustement concerne donc l'ensemble des paramètres du modèle et des caméras et résout, en particulier, le problème de l'estimation de la géométrie inconnue des caméras. En fonction du nombre de données disponibles, dépendant du nombre de paramètres indépendants du modèle et du nombre d'images, il sera donc possible de déterminer une partie, ou l'ensemble, des paramètres des caméras (géométries interne et externe). À noter ici que cette optimisation peut être un processus linéaire ou non-linéaire en fonction de l'application. Dans le deuxième cas, il sera alors souvent nécessaire de fournir des valeurs initiales pour les paramètres à déterminer, ce qui peut s'avérer difficile.

Certaines applications effectuent l'identification et l'ajustement en une seule étape. En effet, un modèle est constitué de contours dont les projections sont caractérisées par de forts contrastes. Une solution consiste donc à optimiser les paramètres du problème de manière à maximiser les contrastes le long des contours projetés dans les images.

1.4.2. *Quelques approches*

Les approches d'ajustement existantes sont nombreuses et variées. Certaines nécessitent une seule image, d'autres plusieurs, par ailleurs l'identification du modèle dans les images peut se faire de manière automatique ou manuelle suivant la méthode employée. Nous présentons ici quelques approches qui illustrent ces différents points.

1.4.2.1. *Une seule image*

Il est possible de modéliser une scène à partir d'une seule image. La contrainte est bien sûr qu'il faut fournir assez d'informations sur la scène pour résoudre les ambiguïtés du passage 2D-3D. En particulier, la géométrie de la caméra, soit la géométrie interne pour une seule caméra, doit être connue ou déterminée. Une approche intéressante repose sur l'ajustement de parallélépipèdes [WIL 01]. Ces derniers sont en

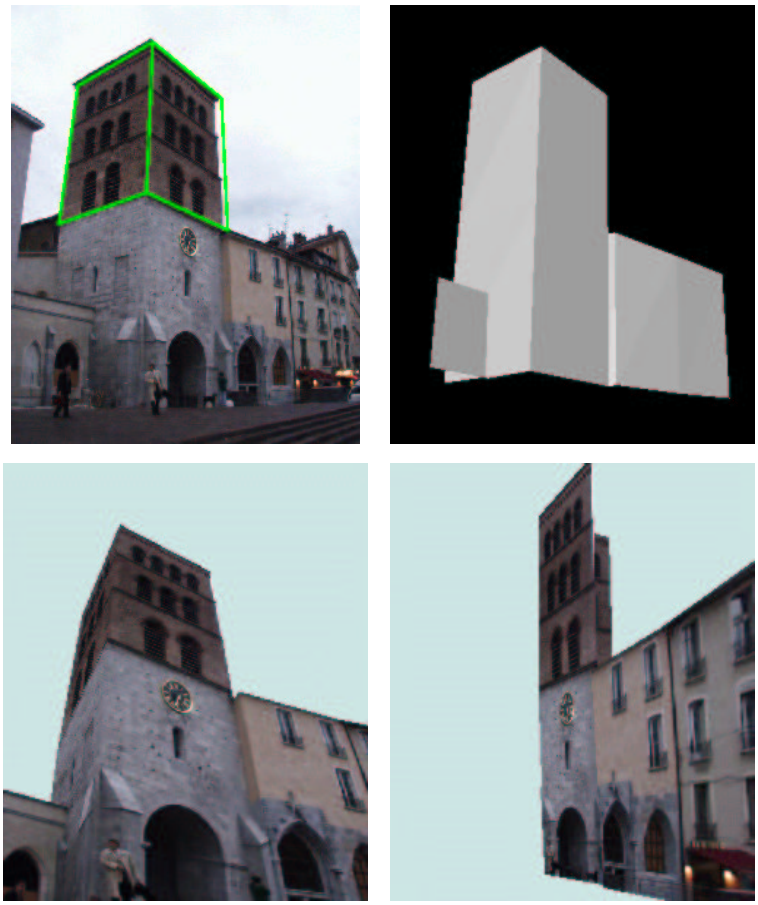

Figure 1.17. *Un exemple de modélisation à partir d'une seule image par ajustement de parallélépipèdes. En haut à gauche l'image ainsi que le parallélépipède utilisés pour déterminer la géométrie interne. En haut à droite le modèle construit (à un facteur d'échelle près) et constitué du parallélépipède ajusté et de plusieurs facettes déterminées à l'aide de contraintes de coplanarité. En bas deux images du modèle texturé (images de M. Wilczkowiak [WIL 01]).*

effet fréquemment présents dans les environnements urbains et permettent de prendre en compte des contraintes de parallélisme et d'orthogonalité. La méthode consiste à définir, dans l'image, la projection d'un, ou de plusieurs, parallélépipèdes. À partir de ces projections et éventuellement de quelques hypothèses faibles sur la caméra (l'axe optique passe au centre de l'image par exemple), il est possible de déterminer l'ensemble de la géométrie interne de la caméra. Cela suffit pour reconstruire le modèle, ici un parallélépipède, à un facteur d'échelle près. À noter que la géométrie externe n'est ici pas nécessaire du fait qu'une seule image soit impliquée.

Le modèle peut ensuite être amélioré en considérant des contraintes appliquées à d'autres primitives, typiquement des points et des segments, extraits manuellement ou automatiquement. Ces contraintes sont la coplanarité ou la colinéarité de plusieurs primitives (voir la figure 1.17).

1.4.2.2. *Plusieurs images et identification manuelle : Facade*

Facade est une approche populaire de ces dernières années [DEB 96]. Le principe est d'utiliser une collection de modèles : parallélépipèdes rectangles, pyramides, dièdres, dont l'utilisateur positionne les projections, à la main, dans toutes les images considérées. Le système ajuste ensuite l'ensemble de ces modèles de manière à ce que leurs projections correspondent à ce que l'utilisateur a défini. Les positions des caméras ainsi que les paramètres des modèles (positions et dimensions) sont déterminés de cette façon. La géométrie interne des caméras est supposée connue et constante à la mise au point près (la distance focale) qui peut elle varier dans les différentes images. Cela suppose donc que l'ensemble des images soient prises avec la même caméra. Cette approche a inspiré un produit commercial : Canoma[9]. La figure 1.18 montre la modélisation du campanile de Berkeley réalisée par cette méthode.

(a) (b) (c)

Figure 1.18. *Le système Facade : (a) l'utilisateur positionne, à la main, les projections des différentes primitives (boîtes, etc.) dans les images ; (b) le système ajuste ensuite ces primitives dans l'espace ; (c) le modèle texturé produit (images de P. Debevec [DEB 96]).*

9. `http ://www.canoma.com/`

1.4.2.3. *Plusieurs images et identification automatique*

L'ajustement automatique de modèles concerne de nombreuses applications dont en particulier : la modélisation de bâtiments (à partir d'images aériennes par exemple), ou la modélisation de visages ou de corps humains. Dans le cas de la modélisation de bâtiments, la plupart des approches supposent qu'une grande partie de la scène peut être modélisée à l'aide de facettes planes. Souvent, des hypothèses, telles que beaucoup de plans de la scène soient verticaux ou horizontaux, sont prises en compte. Une approche classique (voir par exemple [BAI 99]) reconstruit d'abord des primitives individuelles et essaie ensuite d'ajuster des surfaces de types supposés. Les deux classes d'approches, construction et ajustement de modèles, se rejoignent alors (voir la partie 1.3.4.1 page 17).

Des exemples de modèles plus complexes existent pour des bâtiments entiers de différents types (voir par exemple [HEN 96]), des visages ou des corps humains. Une approche modélise par exemple les visages à partir de séquences d'images et d'un modèle initial *générique* de visage [FUA 00]. Cette approche ne nécessite pas d'informations sur les caméras et peu d'interactions humaines : cliquer quelques points spécifiques dans une image. En ce qui concerne le corps humain, beaucoup de recherches sont effectuées sur sa modélisation et sur celle du mouvement. Des systèmes commerciaux existants (les systèmes de *motion capture*) utilisent des marqueurs placés à proximité des jointures (des petites balles ou des vêtements texturés par exemple). Ces marqueurs jouent le même rôle que les cibles dans l'approche photogrammétrique. Après leur reconstruction dans l'espace, un modèle biomécanique (contenant des dimensions mais aussi des contraintes sur les accélérations ou autres) peut être ajusté. Ces systèmes sont couramment utilisés dans l'industrie du cinéma.

Des recherches actuelles tentent de réaliser la même application, sans utilisation de marqueurs. Le problème est alors beaucoup plus difficile, notamment du fait que les primitives ne sont, en général, pas bien définies : les vêtements, par exemple, flottent autour du corps et leurs contours ne constituent pas des primitives très fiables. Quelques-unes des approches existantes utilisent les silhouettes, obtenues par soustraction du fond (voir la partie 1.3.1.3 page 12) ; d'autres prennent en compte la couleur, pour localiser les mains et la tête, ce qui donne des indices sur la position du corps. Enfin certaines couplent plusieurs types d'information dans les images pour ajuster le modèle : les contours qui concernent les parties extrêmes (dans une image) du modèle et les intensités pour les parties intérieures du modèle (voir la figure 1.19)

1.5. Conclusion

Nous avons fait, dans ce chapitre, un tour d'horizon de la modélisation à partir d'images. Nous avons, en particulier, montré les deux grandes classes d'approches qui existent, construction ou ajustement de modèles, en fonction du contexte, inconnu ou

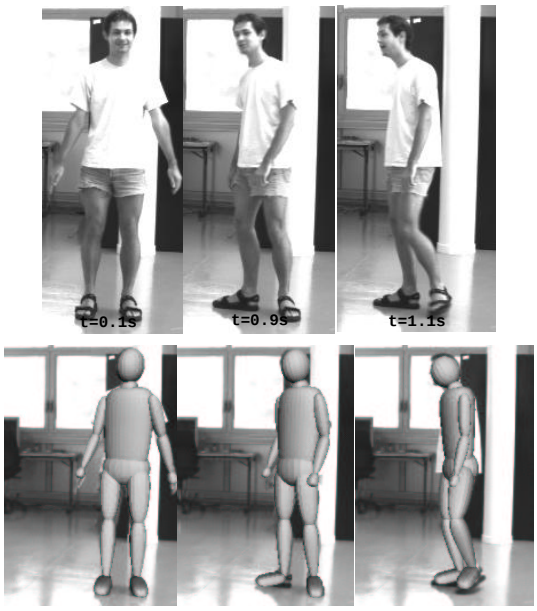

Figure 1.19. *Un exemple de modélisation du corps humain à l'aide de modèles déformables ; ici des super-quadriques. Les informations images utilisées sont les contours ainsi que les variations d'intensités (images de C. Sminchisescu [SMI 01]).*

connu, de l'application. Nous avons, par ailleurs, insisté sur l'importance de la géométrie dans les processus de modélisation. Il est en effet nécessaire de connaître ou de déterminer la géométrie des caméras dans un processus de modélisation. Cela implique la géométrie interne des caméras, soit leurs caractéristiques intrinsèques, ainsi que la géométrie externe des caméras, soit leurs positions relatives. Plusieurs applications de construction ou d'ajustement ont été présentées pour illustrer ces principes. Ces applications démontrent la faisabilité de la modélisation à partir d'images, de nombreux produits commerciaux sont d'ailleurs cités dans le document. Pour conclure ce tour d'horizon, nous identifions quelques problèmes encore *difficiles* à l'heure actuelle en modélisation :

1) Construction de surfaces : dans le cadre de la construction de modèles, la dernière étape consiste à produire une surface à partir des primitives reconstruites : des points ou des segments. Ce problème reste partiellement résolu pour le moment, et en dehors des solutions basées sur des hypothèses sur la surface (planes en grande partie, *etc.*), peu de solutions générales existent.

2) Enveloppe visuelle : les enveloppes visuelles (voir la partie 1.3.5.3 page 21) connaissent actuellement un intérêt important du fait qu'elles conduisent à la reconstruction, approchée, de surfaces, et de ce fait à des applications potentielles en temps réel de réalité virtuelle (pour les studios virtuels par exemple). La difficulté réside dans

la détermination de cette enveloppe, les méthodes actuelles consistant à effectuer des reconstructions approximatives uniquement.

3) Modélisation des parties du corps humain : ce domaine est actif actuellement en recherche, que cela concerne des parties (bras, visages, *etc.*) ou l'ensemble du corps humain. De nombreuses solutions existent, mais très peu fonctionnent en temps réel, et la robustesse reste à améliorer.

4) Automatisation : c'est un aspect un peu plus général que les précédents. Les méthodes de modélisation actuelles requièrent, peu ou prou, l'intervention de l'utilisateur. Une amélioration consisterait à automatiser l'ensemble du processus. Cela implique une robustesse sans faille des différentes étapes du processus de modélisation, ce qui est toujours difficile à obtenir.

À cela s'ajoutent, bien entendu, d'autres problèmes plus généraux en vision par ordinateur, telle que la mise en correspondance, le suivi automatique de primitives, *etc.*, qui restent difficiles à ce jour.

1.6. Bibliographie

[AME 98] AMENTA N., BERN M., KAMVYSSELIS M., « A New Voronoi-Based Surface Reconstruction Algorithm », *ACM Computer Graphics (Proceedings SIGGRAPH)*, p. 415–421, 1998.

[BAI 99] BAILLARD C., ZISSERMAN A., « Automatic Reconstruction of Piecewise Planar Models from Multiple Views », *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, IEEE Computer Society Press, Wed May 17 2000, p. 559–565, June 1999.

[BOY 00] BOYER E., PETITJEAN S., « Curve and Surface Reconstruction From Regular and Non-Regular Point Sets », *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, (USA)*, vol. II, p. 659–665, 2000.

[COC 95] COCQUEREZ J., PHILIP S., *Analyse d'images : Filtrage et Segmentation*, Masson. Paris, 1995.

[CUR 96] CURLESS B., LEVOY M., « A volumetric method for building complex models from range images », *ACM Computer Graphics (Proceedings SIGGRAPH)*, p. 303–312, 1996.

[DEB 96] DEBEVEC P., TAYLOR C., MALIK J., « Modeling and Rendering Architecture from Photographs : A Hybrid Geometry- and Image-Based Approach », *ACM Computer Graphics (Proceedings SIGGRAPH)*, p. 11–20, 1996.

[FAU 87] FAUGERAS O., TOSCANI G., « Camera Calibration for 3D Computer Vision », *Proceedings of International Workshop on Machine Vision and Machine Intelligence, Tokyo, Japan*, 1987.

[FAU 92] FAUGERAS O., « What Can Be Seen in Three Dimensions with An Uncalibrated Stereo Rig ? », SANDINI G., Ed., *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, Springer-Verlag, Fri Jan 10 1992, p. 563–578,

May 1992.

[FAU 93] FAUGERAS O., *Three-Dimensional Computer Vision : A Geometric Viewpoint*, Artificial Intelligence, MIT Press, Cambridge, 1993.

[FAU 01] FAUGERAS O., LUONG Q.-T., PAPADOPOULO T., *The Geometry of Multiple Images*, MIT Press, March 2001.

[FIT 98] FITZGIBBON A., CROSS G., ZISSERMAN A., « Automatic 3D Model Construction for Turn-Table Sequences », KOCH R., VANGOOL L., Eds., *Proceedings of SMILE Workshop on Structure from Multiple Images in Large Scale Environments*, vol. 1506 de *Lecture Notes in Computer Science*, Springer Verlag, p. 154-170, June 1998.

[FUA 00] FUA P., « Regularized Bundle-Adjustment to Model Heads from Image Sequences without Calibration Data », *International Journal of Computer Vision*, vol. 38, n°2, p. 153–171, 2000.

[FUS 00] FUSIELLO A., « Uncalibrated Euclidean reconstruction : A review », *Image and Vision Computing*, vol. 18, n°6-7, p. 555–563, 2000.

[HAR 88] HARRIS C., STEPHENS M., « A combined corner and edge », *Proceedings 4th Alvey Vision Conference*, p. 147–151, 1988.

[HAR 92] HARTLEY R., GUPTA R., CHANG T., « Stereo from Uncalibrated Cameras », *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, Mon Sep 7 1992, p. 761–764, 1992.

[HAR 95] HARTLEY R., « In Defence of the 8-points Algorithm », *Proceedings of 5th International Conference on Computer Vision, Boston (USA)*, p. 1064–1070, janvier 1995.

[HAR 97] HARTLEY R., STURM P., « Triangulation », *Computer Vision and Image Understanding*, vol. 68, n°2, p. 146–157, 1997.

[HAR 00] HARTLEY R., ZISSERMAN A., *Multiple View Geometry in Computer Vision*, Cambridge University Press, juin 2000.

[HEN 96] HENRICSSON O., GRUEN A., « Overview of Research Activities at ETH-Zürich in Automated 3-D Reconstruction of Buildings from Aerial Images », *Tagungsband der 16. Wissenschaftlich-Technischen Jahrestagung der DGPF*, n° 5Publikationen der DGPF, Deutsche Gesellschaft für Photogrammetrie und Fernerkundung, Thu Feb 27 1997, 1996.

[HOP 92] HOPPE H., DEROSE T., DUCHAMP T., MCDONALD J., STUETZLE W., « Surface Reconstruction from Unorganized Points », *ACM Computer Graphics (Proceedings SIGGRAPH)*, vol. 26(2), p. 71–78, juillet 1992.

[HOR 99] HORPRASERT T., HARWOOD D., DAVIS L., « A Statistical Approach for Real-time Robust Background Subtraction and Shadow Detection », *IEEE ICCV'99 FRAME-RATE WORKSHOP*, 1999.

[KUT 00a] KUTULAKOS K., « Approximate N-View Stereo », *Proceedings, 6th European Conference on Computer Vision, Dublin, (Ireland)*, p. 67–83, 2000.

[KUT 00b] KUTULAKOS K., SEITZ S., « A Theory of Shape by Space Carving », *International Journal of Computer Vision*, vol. 38, n°3, p. 199–218, 2000.

[LAU 94] LAURENTINI A., « The Visual Hull Concept for Silhouette-Based Image Understanding », *IEEE Transactions on PAMI*, vol. 16, n° 2, p. 150-162, février 1994.

[LAZ 01] LAZEBNIK S., BOYER E., PONCE J., « On How to Compute Exact Visual Hulls of Object Bounded by Smooth Surfaces », *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Kauai, (USA)*, vol. I, p. 151-161, December 2001.

[LUO 96] LUONG Q., FAUGERAS O., « The Fundamental Matrix : Theory, Algorithms and Stability Analysis », *International Journal of Computer Vision*, vol. 17, n° 1, p. 43–75, 1996.

[MAT 01] MATUSIK W., BUEHLER C., MCMILLAN L., « Polyhedral Visual Hulls for Real-Time Rendering », *Eurographics Workshop on Rendering*, 2001.

[MOR 00] MORRIS D., KANADE T., « Image-Consistent Surface Triangulation », *Proceedings of the Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, 2000.

[NIE 94] NIEM W., BUSCHMANN R., « Automatic Modelling of 3D Natural Objects from Multiple Views », *European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production, Hamburg, Germany*, Wed Jan 29 1997, November 1994.

[POL 98] POLLEFEYS M., KOCH R., VAN GOOL L., « Self-Calibration and Metric Reconstruction in Spite of Varying and Unknown Internal Camera Parameters », *Proceedings of the 6th International Conference on Computer Vision, Bombay, (India)*, Wed Aug 27 1997, p. 90–95, 1998.

[PRA 78] PRATT W., *Digital Image Processing*, Wiley-Interscience, 1978.

[SCH 98] SCHMID C., MOHR R., BAUCKHAGE C., « Comparing and Evaluating Interest Points », *Proceedings of the 6th International Conference on Computer Vision, Bombay, (India)*, Fri May 16 1997, p. 230–235, 1998.

[SCH 00] SCHMID C., ZISSERMAN A., « The geometry and matching of lines and curves over multiple views », *International Journal of Computer Vision*, vol. 40, n° 3, p. 199–234, 2000.

[SHA 95] SHASHUA A., « Algebraic Functions for Recognition », IEEE *Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, n° 8, p. 779–789, August 1995.

[SLA 80] SLAMA C., Ed., *Manual of Photogrammetry, Fourth Edition*, American Society of Photogrammetry and Remote Sensing, Falls Church, Virginia, USA, 1980.

[SMI 01] SMINCHISESCU C., TRIGGS B., « Covariance Scaled Sampling for Monocular 3D Human Tracking », *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Kauai, (USA)*, 2001.

[STU 99] STURM P., MAYBANK S., « On Plane-Based Camera Calibration : A General Algorithm, Sngularities, Applications », *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, (USA)*, p. 432–437, juin 1999.

[SUL 98] SULLIVAN S., PONCE J., « Automatic Model Construction, Pose Estimation, and Object Recognition from Photographs Using Triangular Splines », *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, Mon Jul 20 1998, p. 510–516, January 1998.

[SZE 93] SZELISKI R., « Rapid Octree Construction from Image Sequences », *Computer Vision, Graphics and Image Processing*, vol. 58, n° 1, p. 23–32, July 1993.

[TOY 99] TOYAMA K., KRUMM J., BRUMITT B., MEYERS B., « Wallflower : Principles and Practice of Background Maintenance », *Proceedings of the 7th International Conference on Computer Vision, Kerkyra, (Greece)*, p. 255–261, 1999.

[TRI 95] TRIGGS B., « Matching Constraints and the Joint Image », *Proceedings of 5th International Conference on Computer Vision, Boston (USA)*, p. 338–343, juin 1995.

[TRI 99] TRIGGS B., MCLAUCHLAN P. F., HARTLEY R. I., FITZGIBBON A. W., « Bundle Adjustment — A Modern Synthesis », TRIGGS B., ZISSERMAN A., SZELISKI R., Eds., *Vision Algorithms : Theory and Practice*, n° 1883LNCS, Corfu, Greece, Springer-Verlag, p. 298–373, septembre 1999.

[TSA 87] TSAI R., « A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses », *IEEE Journal of Robotics and Automation*, vol. RA-3, n° 4, p. 323-344, 1987.

[WIL 01] WILCZKOWIAK M., BOYER E., STURM P., « Camera Calibration and 3D Reconstruction from Single Images Using Parallelepipeds », *Proceedings of the 8th International Conference on Computer Vision, Vancouver, (Canada)*, vol. I, p. 142–148, July 2001.